THESIS / REPORTS
STEELE, B.M.

Final Report for 03-JV-11222048151 (LANDFIRE STATPAK)

Statistical Methods to Evaluate Departure from Reference Conditions

Steele, B.M. Principal Investigator, Reddy, S.K., Research Associate[1]

September 30, 2004

1. Introduction

Many forests and rangelands of North America experienced frequent, low severity fires prior to European settlement (circa 1900) (Wright and Heinselman 1973, Heinselman 1981, DeBano et al. 1998). However, land management policies of the past century, such as fire exclusion, livestock grazing, and logging, coupled with exotic species invasions, have altered the fuel loadings, vegetation composition and structure, and consequently, the frequency and severity of fire in these ecosystems (Ferry et al. 1995, Frost 1998, Kolb 1998, Arno et al. 2000, Keane et al. 2002b). Dramatic increases in the number, size, and intensity of wildfires in recent years are partially attributed to these changes (GAO 2002). In response, Congress and the Executive Branch mandated the development of a National Fire Plan (www.fireplan.gov) and directed the Departments of Agriculture and Interior to develop a cohesive strategy for implementing the Plan (Laverty and Williams 2000). These efforts, along with the passage of the Healthly Forests Restoration Act (www.healthyforests.gov), provide support for reducing fire hazard using silvicultural thinning and prescribed burning.

The Cohesive Strategy uses fire regime condition class (FRCC) as a key metric for implementating the National Fire Plan. FRCC is based on the departure of current conditions from historical conditions described as changes in fuel, fire regime, and vegetation characteristics (Laverty and Williams 2000). This metric is used to allocate funding and resources for those lands with the greatest need for restoration as directed in the National Fire Plan. However, data on historical conditions is extremely sparse (Keane et al. 2002a). Several strategies have been used to quantify historical conditions (Landres et al. 1999), but the strategy most amenable for large-scale application generates historical data by simulating plant community succession (Keane et al. 2002a).

Using simulated data is no longer unusual in ecology and resource management, and is become common-plce for large-scale problems (Waring and Running 1998). Succession and fire modeling has been an active area of ecological research for three decades (see Keane et al, 2004 and Mladendoff and Baker 1999); and these models are now mature in their sophistication and accuracy. Limitations related to simulation model assumptions and accuracy are widely recognized (Botkin and Schenk 1996, Keane et al. 2001). Regrettably, questions regarding the accuracy of these model generally cannot be subjected to effective statistical analysis without extensive observational data. Despite the limitations and questions regarding accuracy, simulation is currently the only viable approach for assessing large areas lacking adequate historical chronosequences. However, the analysis of simulated ecological data pose statistical challenges because data sets are often massive, and consist of observations that are

[1]Keane, R.F. USDA FS, Rocky Mountain Res. Station contributed to parts of this report

emailed to G&A 10-22-04

Thesis/
Reports
Steele,
B. M.

high-dimensional or autocorrelated, or both. This article presents a methodology for comparing simulated, spatially explicit historical chronosequences with current conditions. The methodology is based on a departure index suitable for mapping and identifying lands that require fuel treatment and ecosystem restoration under the direction of the National Fire Plan's Cohesive Strategy.

U.S. Federal agencies charged with the management of public lands have adopted a two-tier hierarchical plant community classification system (citations; Arno et al. 1985?). These systems are developed by analyzing the existing plant communities over large landscapes such as the interior Columbia Basin (Hessburg, et al. 1999; Hann et al. 1998). The first tier, potential vegetation type (PVT), describes the biophysical conditions at a site by identifying the plant community that would exist if succession were to progress to the stable or climax community. Within a PVT, the timing and intensity of disturbance events interrupt or set back succession. After disturbance, progression towards succession proceeds along a pathway that may differ from original depending on the structure of the surviving community. For example, a low intensity fire may leave a mulitstrata tree canopy and a high intensity fire may eliminate the forest canopy entirely, and so the successional communities that develop after fire will differ. The second tier of the classification system consists of the successional classes that develop along these pathways. Successional class describes the overstory structure and the dominant overstory species (e.g., subalpine fir). In the absence of disturbance, transition between successional classes is viewed as a deterministic process governed by species-specific environmental requirements, inter-species competition, and environmental conditions at the site. Figure 1 shows an example of the successional pathways for a PV type occurring in the Forest ecosystems of Utah.

1.1. Background

Federal land management agencies use a two-tier hierarchical plant community classification system for describing landscapes such as the existing plant communities of the interior Columbia Basin (Hessburg, et al. 1999; Hann et al. 1998). The first tier, potential vegetation type (PVT), describes the biophysical conditions at a site by identifying the plant community that would exist if succession were to progress to the stable or climax community (Daubenmire 1966, Pfister and Arno 1980). Within a PVT, the timing and intensity of disturbance events, especially wildland fire, interrupt or set back succession (Arno et al. 1985). After disturbance, progression towards succession proceeds along a pathway whose trajectory is governed by the structure of the surviving community and the disturbance severity. For example, a low intensity fire may kill many understory trees leaving a single-strata overstory tree canopy while a high intensity fire may kill all trees. The second tier of the classification system consists of the successional classes that develop along these pathways. Successional class describes the vertical stand structure and dominant overstory species (e.g., dense subalpine fir). In the absence of disturbance, transition between successional classes is viewed as a deterministic process governed by species-specific environmental requirements, inter-species competition, and environmental conditions at the site.

The focus of this article is on comparing a single observation on the current, or observed distribution of successional classes to the simulated historical distribution for each stratum within

a management zone. These data pose several challenges. Because the data are observations on high-dimensional multinomial categorical variables dominated by small values (≤ 2) classical, distribution-based statistical analysis is problematic. Moreover, the data matrix $\boldsymbol{X}$ is not full rank, which complicates the use of some multivariate methods . These considerations lead to us to using a singular value decomposition to obtain a full rank approximation of $\boldsymbol{X}$. The departure of an observation vector $\boldsymbol{x}_0$ from historic conditions is measured by the length of the error vector obtained from projecting $\boldsymbol{x}_0$ onto the linear space spanned by $\boldsymbol{X}$. To determine whether this departure is significant, an empirical distribution of departure values is obtained by drawing each historic observation $\boldsymbol{x}_i$ from $\boldsymbol{X}$ without replacement, and computing the departure of $\boldsymbol{x}_i$ from the remaining data set. The comparison of current to historic departures provides a measure of the strength of evidence of departure. In addition to the dimensionality problem, the absence of appropriate distributional models are among the recent challenges facing statisticians as a consequence of technological innovations (Rao 2004). In addition, this methodology was developed to be applied over very large areas primarily by automated procedures using massive data sets. These types of large-scale analyses are becoming common in ecology and resource managment, and demand statistical methods that are both sensitive and computationally efficient.

Before proceeding further, it is convenient to establish notation and terminology.

1.2. Notation and Terminology

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ denote n observations on p variables that have been sampled from a stochastic process $\mathcal{S}$. Significance testing, but not estimation of departure requires the assumptions that the observations are independent. For now, we proceed without assuming independence. The observation $\boldsymbol{x}_0$ is an additional observation on the same variables, and in particular, $\boldsymbol{x}_0$ is an observation on current conditions. For simplicity, we use the notation $\boldsymbol{x}_0 \in \mathcal{S}$ and $\boldsymbol{x}_0 \notin \mathcal{S}$ to state that $\boldsymbol{x}_0$ was generated by $\mathcal{S}$, and not generated by $\mathcal{S}$, respectively. The $\mathcal{L}_2$ norm of $\boldsymbol{x}_i$ is denoted as $\|\boldsymbol{x}_i\|_2$. The $p \times p$ identity matrix is denoted by $\boldsymbol{I}_p$, $\mathbf{1}$ denotes a p-vector of ones, and $\#A$ denotes the cardinality of the set A. The data are collected as an $n+1 \times p$ matrix

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_0^T \\ \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix},$$

where $\boldsymbol{x}_i^T$ denotes the transpose of $\boldsymbol{x}_i$. The i,jth element of $\boldsymbol{X}$ is denoted by $\boldsymbol{x}_{ij}$, and $\overline{\boldsymbol{x}}_{\cdot j}$ $= n^{-1}\sum_{i=1}^{n} \boldsymbol{x}_{ij}$ denotes the mean of the ith column. The p-vector of column means is $\overline{\boldsymbol{x}}$ $= (\overline{\boldsymbol{x}}_{\cdot 1}, \ldots, \overline{\boldsymbol{x}}_{\cdot p})^T = n^{-1}\boldsymbol{X}^T\mathbf{1}$. The $p \times p$ sample variance matrix is $\boldsymbol{D} = n^{-1}\boldsymbol{X}^T(\boldsymbol{I}_p$ $- n^{-1}\mathbf{1}\mathbf{1}^T)\boldsymbol{X}$. The spectral decomposition of $\boldsymbol{D}$ is $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^T$ where $\boldsymbol{\Lambda}$ $=$ diag$(\lambda_1 \ \lambda_2 \ \cdots \ \lambda_p)$ is a diagonal matrix comprised of the eigenvalues of $\boldsymbol{D}$ with the eigenvalues arranged in descending order i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. The matrix $\boldsymbol{A}$ $= (\boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \cdots \ \boldsymbol{a}_p)$ is comprised of the normalized eigenvectors of $\boldsymbol{D}$, and the ith principal component isdenoted by $\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{a}_i$. For the problem at hand, $\boldsymbol{X}$ is not full rank and $\boldsymbol{D}$ is not

invertible because each PV type imposes a linear constraint on the columns of $\boldsymbol{X}$. The Moore-Penrose generalized inverse of $\boldsymbol{D}$ can be computed as

$$\boldsymbol{D}^{-} = \boldsymbol{V}_r \boldsymbol{\Lambda}_r^{-1} \boldsymbol{V}_r^T$$

(Schott 1997, Chap. 5), where r is the rank of $\boldsymbol{X}$. Because $\lambda_i > 0$ for $i = 1, \ldots, r$, $\boldsymbol{D}^{-}$ is positive definite and can be used to define a vector norm $\|\boldsymbol{x}_0\|_{\boldsymbol{D}^{-}} = \sqrt{\boldsymbol{x}_0^T \boldsymbol{D}^{-} \boldsymbol{x}_0}$ and a distance function $\rho(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{D}^{-}}$. Finally, the matrix constructed by removing the ith row, or observation, from $\boldsymbol{X}$ is denoted by $\boldsymbol{X}_i$.

2. Methodology

The methodology for assessing departure of current conditions from historical consists of two distinct stages, generating data by simulating of plant community succession under historic fire regimes, and the comparison of the observed plant community distribution to the simulated historic data.

2.2. Simulating Plant Community Succession

The LANDscape SUccession Model (LANDSUM) (Keane et al. 2002) was used to generate observations on successional class distribution. The predecessor of LANDSUM is CRBSUM, a model used to simulate coarse scale vegetation dynamics in the interior Columbia River Basin at a 1 km pixel resolution (Keane et al. 1996). Unlike CRBSUM, LANDSUM model is spatially explicit. Simulations are carried out for each 30 m × 30 m pixel comprising a lattice covering a region of 100 to 10,000 km^2. A geographic information system maintains the lattice and stores values on topography, PVT and prevailing wind direction for each pixel. From an initial state, time is simultaneously incremented for all pixels. As time advances, a time series of successional classes is generated for each pixel in response to succession and fire disturbance. Successional transition between classes occurs at pre-determined time intervals along a PVT-specific successional pathway, as shown in Figure 1. In contrast, fire events are stochastically generated according to PVT-specific ignition probabilities, and fire is spread stochastically from pixel to pixel as a function of prevailing wind direction and speed and slope (Keane et al. 2002; Rothermel 1991). Fire events modify successional trajectory according to fire intensity (low, medium or high intensity) (Keane et al. 1996).

A simulation is carried out for a core region (the area of interest) by enlarging the simulation region to include a perimeter region. The perimeter allows for the spread of fires into the core region from outside. The initial state for each pixel is defined by the modal succession class for the pixel's previously identified PV type, and time is incremented in one-year steps for 5,000 years. Though the simulation unit is a pixel, a coarser lattice consisting of 900×900 m strata (each stratum contains 900 pixels) is used to compare current and historic conditions. The coarse lattice is used because it conforms with the GIS lattice used for management decision-making. Simulated successional processes are remarkably different even among adjacent strata because of spatial variation in PVT distribution and fire spread characteristics induced by variation in terrain. Because spatial correlation tends to be minor and inconsistent, neighboring strata are ignored in the analysis of each stratum.

Output from the first 500 years of the simulation are discarded, and data from the remaining years are sampled every 20 years for the comparison of current and historical conditions. This initialization strategy was adopted after comparing each year to a set of comprised of 10,000 years using the departure measure discussed below. This analysis showed that the most of the transition from the initial state of modal successional classes to a stationary process was complete after 400 years. With respect to serial correlation, a 50 year sampling interval appears sufficient for producing a set of departure values free of serial correlation. However, we chose to sample on 20 year intervals to limit the computational cost of long simulation times and collect data for 4000 years after the initialization period. The interpretation of comparisons of current and historical conditions given that the simulated historical observations are not independent is discussed below. It should be noted that ecological regions besides central Utah forests may differ with respect and autocorrelation and initialization period.

2.3. Comparing Observed and Simulated Plant Community Distributions

In general terms, the objective is to determine whether anobservation $\boldsymbol{x}_0$ describing the current, or existing, distribution of successional classes for a stratum is unusual with respect to the stochastic process that generated the n observation vectors $\boldsymbol{x}_i$ on historical conditions. The data matrix $\boldsymbol{X}$ is comprised of the $n+1$ row vectors $\boldsymbol{x}_i$, $i = 0, 1, \ldots, n$, and each row vector consists of the count of pixels belonging all PVT/successional classes encountered in the stratum simulation, or under current conditions. For the central Utah data, the number of encountered successional classes p varied between approximately 50 and 250. A column constraint is imposed on $\boldsymbol{X}$ by the condition that the number of pixels is fixed at 900 for each year. Each PVT induces an additional linear constraint on the columns of $\boldsymbol{X}$ because pixel PVT is static; hence, the total number of pixels distributed across successional classes within a PVT is constant

There are similarities between this problem and two-sample and multivariate outlier detection problems. With respect to the two-sample approach, it is useful to view the data matrix as a contingency table with PVT/successional class as columns and $n+1$ years as rows. Typically, the majority of the pixel counts are 0 and, correspondingly, the sample mean successional class frequency $\overline{\boldsymbol{x}}_{\cdot j}$ is less than 1 for most successional classes. Given that most counts are small and that the observations $\boldsymbol{x}_i$, $i = 1, \ldots, n$, exhibit some serial correlation, it is inappropriate to model the data using the multinomial distribution and to assess departure by conducting a chi-square test of homogeniety of proportions. On the other hand, it may be useful to investigate a measure of distance between $\boldsymbol{x}_0$ and $\boldsymbol{X}$ that is analogous to the chi-square statistic for a test of homogeneity of proportions. One such measure is the distance $d_\chi(\boldsymbol{x}_0) = \left\| \underset{p \times p}{\boldsymbol{W}} (\boldsymbol{x}_0 - \overline{\boldsymbol{x}}) \right\|_2$ where $\boldsymbol{W}$ is diagonal with $\boldsymbol{W}_{jj} = \overline{\boldsymbol{x}}_{\cdot j}^{-1/2}$ if $\overline{\boldsymbol{x}}_{\cdot j} \geq 3$ and $\boldsymbol{W}_{jj} = 0$ if $\overline{\boldsymbol{x}}_{\cdot j} < 3$. The squared distance $d_\chi^2(\boldsymbol{x}_0)$ is an approximation (though perhaps poor) of the usual chi-square statistic $\sum \overline{\boldsymbol{x}}_{\cdot j}^{-1} (\boldsymbol{x}_{0j} - \overline{\boldsymbol{x}}_{\cdot j})^2$ that avoids small counts. Instead of using the chi-square distribution to approximate observed significance level of a test of $H_0 : \boldsymbol{x}_0 \in \mathcal{S}$ versus $H_1 : \boldsymbol{x}_0 \notin \mathcal{S}$, we compute the proportion of distances $d_\chi(\boldsymbol{x}_i)$, $i = 0, 1, \ldots, n$ less or equal to $d_\chi(\boldsymbol{x}_0)$. Because row vectors are not necessarily independent, this proportion should not be interpreted

as the observed significance level, but rather a comparative measure of evidence. As such, it is useful for comparing among different strata to identify those strata with the greatest evidence of departure.

Assessing departure is similar to identifying multivariate outliers as both tasks attempt to identify atypical observations by comparing a candidate observation to a larger set of observations. Multivariate outlier detection methods tend to be largely distribution-free and suited for high dimensional data; these characteristics are important for departure assessment. Traditional multivariate outlier detection approaches identify observations that are distant from a measure of central tendency, usually the sample mean. In addition, distance measures are commonly based on the spectral decomposition of the sample variance matrix $\boldsymbol{D}$ [see Jolliffe (2002), Chap. 12]. For example, Mahalanobis distance, the simplest and perhaps the most commonly used multivariate outlier detection measure is $d_M(\boldsymbol{x}_0)$ $= \sqrt{(\boldsymbol{x}_0 - \overline{\boldsymbol{x}})^T \boldsymbol{D}^{-1} (\boldsymbol{x}_0 - \overline{\boldsymbol{x}})}$, where $\boldsymbol{D}$ is assumed to be full rank. When $\boldsymbol{D}$ is not full rank, then an analogous distance measure is $d_M^r(\boldsymbol{x}_0) = \left\| \boldsymbol{\Lambda}_r^{-1/2} \boldsymbol{V}_r (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \right\|_2$. The distance measure d_M^r gives equal weight to each eigenvector axis in the sense that the sample variance of the vector $\boldsymbol{\Lambda}_r^{-1/2} \boldsymbol{V}_r (\boldsymbol{x}_i - \overline{\boldsymbol{x}})$ is $\boldsymbol{I}_r$. Gnanadesikhan and Kettenring (1972) proposed an alternate distance measure that emphasizes those variables with larger sample variances. When $\boldsymbol{D}$ is not full rank, then the measure can be expressed as $d_{GK}^r(\boldsymbol{x}_0) = \left\| \boldsymbol{\Lambda}_r^{1/2} \boldsymbol{V}_r (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \right\|_2$. Observations that are distant with respect to d_M^r and d_{GK} tend to be outliers respect to one or more of the univariate variables, and are often detectable by univariate methods. Consequently, other multivariate outlier detection methods usually attempt to expose observations that are unusual in some respect that is not obvious using univariate methods (Jolliffe 2002, Chap. 10). For example, if $\boldsymbol{D}$ is full rank, and $\boldsymbol{V}_s$ denotes a matrix constructed from those eigenvectors corresponding to s smallest eigenvalues, then $d_1^s(\boldsymbol{x}_0) = \left\| \boldsymbol{V}_s (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \right\|_2$ and $d_2^s(\boldsymbol{x}_0)$ $= \left\| \boldsymbol{\Lambda}_s^{-1/2} \boldsymbol{V}_s (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \right\|_2$ can be used of outlier detection. These schemes place greater weight on those eigenvectors with the least contribution towards explaining variation in $\boldsymbol{X}$. However, there may be problems in discriminating between very small non-negative eigenvalues and zero-valued eigenvalues when $\boldsymbol{X}$ is not full rank because of computer precision limitations.

There are other limitations associated with these multivariate outlier detection approaches for the problem at hand. Most importantly, the process S may generate several modes, and measuring distance to a single centroid may identify observations as outliers that are near a mode. In addition, the appropriate number of eigenvector axes with which to measure distance is difficult to determine in advance, and hence poses problems if the method is to be applied without human intervention. More recent outlier detection methods attempt to resist preserve sensitivity when many outliers are present, sometimes called robustness against swamping (Jackson and Chen 2004; Juan and Preito 2001; Penny and Jolliffe 2001; Pena and Preito 2001;). The computational demands of these methods are prohibitive for this application in which departure must be evaluated for millions of strata; moreover, swamping is not a concern because the simulated historic data is free of observational error.

2.4. The Row Projection Measure of Departure

We propose a new approach based the extent to which a suspected outlier $\boldsymbol{x}_0$ can be approximated as a linear combination of the observations comprising $\boldsymbol{X}_0$. More specifically, the method projects the normalized vector $\boldsymbol{x}_0$ onto the space spanned by the rows of $\boldsymbol{X}_0$. The departure of $\boldsymbol{x}_0$ is the length of the residual, or error vector. If departure is unusually large, then $\boldsymbol{x}_0$ is said to be atypical of the process that generated $\boldsymbol{X}_0$. Let $\boldsymbol{z}_0$ denote the normalized version of test vector $\boldsymbol{x}_0$, i.e., $\boldsymbol{z}_0 = \boldsymbol{x}_0 / \|\boldsymbol{x}_0\|_2$. Let $\boldsymbol{P}_0$ denote a projection matrix for $\mathcal{R}(\boldsymbol{X}_0)$, the linear subspace of $\mathbb{R}^p$ spanned by the rows of $\boldsymbol{X}_0$. If $\boldsymbol{X}_0^T$ is full rank, then

$$\underset{p\times p}{\boldsymbol{P}_0} = \boldsymbol{X}_0^T(\boldsymbol{X}_0\boldsymbol{X}_0^T)^{-1}\boldsymbol{X}_0. \tag{1}$$

A projection matrix for the null space of $\mathcal{R}(\boldsymbol{X}_0)$ is $\boldsymbol{P}_0^{\perp} = \boldsymbol{I}_p - \boldsymbol{P}_0$, and the projection of $\boldsymbol{x}_0$ onto the null space is $\boldsymbol{v}_0 = \boldsymbol{P}_0^{\perp}\boldsymbol{z}_0$. Our distance measure between $\boldsymbol{x}_0$ and $\boldsymbol{X}_0$ is the length of the error vector $\|\boldsymbol{v}_0\|_2$. Schott (1997, Chap. 2) provides an extended discussion of projection matrices and linear spaces.

The projection $\boldsymbol{u}_0$ is a linear approximation of $\boldsymbol{z}_0$ constructed from the rows of $\boldsymbol{X}_0$ with the least possible squared error $\|\boldsymbol{v}_0\|_2^2$. The projection is also the least squares prediction of $\boldsymbol{z}_0$ obtained by regressing $\boldsymbol{z}_0$ on the rows of $\boldsymbol{X}_0$. In addition, $\|\boldsymbol{v}_0\|_2$ is related to the proportion of variation in $\boldsymbol{z}_0$ not explained by design matrix $\boldsymbol{X}_0^T$. In this context, the sum-of-squares due to the regression is $S = \boldsymbol{z}_0^T\boldsymbol{P}_0\boldsymbol{z}_0 - n\bar{\boldsymbol{z}}_0^2$, and the coefficient of determination is $R^2 = S/(1 - n\bar{\boldsymbol{z}}_0^2)$. Thus, the proportion of variation in $\boldsymbol{z}_0$ not explained by the rows of $\boldsymbol{X}_0$ is

$$1 - R^2 = \frac{1 - \|\boldsymbol{v}_0\|^2}{1 - n\bar{\boldsymbol{z}}_0^2}.$$

For the problem at hand, the $n \times n$ matrix $\boldsymbol{X}_0\boldsymbol{X}_0^T$ is not full rank and formula (1) cannot be used to compute the projection matrix. The singular value decomposition of $\boldsymbol{X}^T$ provides an alternative. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the ordered eigenvalues of $\boldsymbol{X}\boldsymbol{X}^T$ (the first p of which are also the eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$) and $\boldsymbol{\Lambda}$ denote the diagonal matrix with diagonal elements $\sqrt{\lambda_k}$, $k = 1, \ldots, n$. The singular value decomposition of $\boldsymbol{X}^T$ can be expressed as

$$\underset{p\times n}{\boldsymbol{X}^T} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T,$$

where $\underset{p\times p}{\boldsymbol{U}}$ and $\underset{n\times n}{\boldsymbol{V}}$ are comprised of the eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{X}\boldsymbol{X}^T$, respectively. Suppose that the rank of $\boldsymbol{X}^T$ is $r \leq n$ and $\boldsymbol{\Lambda}_r$ denotes the square submatrix of $\boldsymbol{\Lambda}$ containing

the $n - r$ nonsingular eigenvalues. The truncated singular value decomposition of $\boldsymbol{X}^T$ is

$$\begin{aligned}\boldsymbol{X}^T &= \boldsymbol{U}\begin{pmatrix}\boldsymbol{\Lambda}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{V}^T\\ &= (\boldsymbol{U}_r \;\; \boldsymbol{U}_{n-r})\begin{pmatrix}\boldsymbol{\Lambda}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}_{n-r}\end{pmatrix}\begin{pmatrix}\boldsymbol{V}_r^T\\ \boldsymbol{V}_{n-r}^T\end{pmatrix}\\ &= \boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^T\end{aligned}$$

(Vogel 2002, Chap. 1). A generalized inverse of $\boldsymbol{X}\boldsymbol{X}^T$ is $(\boldsymbol{X}\boldsymbol{X}^T)^- = \boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^T$,

$$\begin{aligned}(\boldsymbol{X}\boldsymbol{X}^T)^- &= \boldsymbol{V}\boldsymbol{\Lambda}^{-2}\boldsymbol{V}^T\\ &= \boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^T,\end{aligned}$$

where $\boldsymbol{\Lambda}^{-2}$ is defined to be the diagonal matrix with diagonal elements λ_i^{-2} if $\lambda_i > 0$ and 0 if $\lambda_i = 0, i = 1, \ldots, n$, and $\boldsymbol{V}_r$ and $\boldsymbol{\Lambda}_r^{-2}$ are the $n \times r$ and $r \times r$ submatrices of $\boldsymbol{V}$ and $\boldsymbol{\Lambda}^{-2}$ corresponding to the non-singular eigenvalues. Substituting $\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^T$ for $\boldsymbol{X}_0^T$ and $\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^T$ for $(\boldsymbol{X}_0\boldsymbol{X}_0^T)^{-1}$ in formula (1) leads to

$$\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^T\boldsymbol{V}_r\boldsymbol{\Lambda}_r^{-2}\boldsymbol{V}_r^T\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^T = \boldsymbol{U}_r\boldsymbol{U}_r^T.$$

It is straightforward to verify that $\boldsymbol{P}_0^{\perp} = \boldsymbol{I}_p - \boldsymbol{U}_r\boldsymbol{U}_r^T$ is the projection matrix for the null space of $\mathcal{R}(\boldsymbol{X}_0)$ (Schott 1997, Chap. 2) and that the norm of the error vector $\boldsymbol{v}_0 = \boldsymbol{P}_0^{\perp}\boldsymbol{z}_0$ is $\|\boldsymbol{v}_0\|_2 = \sqrt{1 - \boldsymbol{z}^T\boldsymbol{U}_r\boldsymbol{U}_r^T\boldsymbol{x}_0}$. We define the departure of $\boldsymbol{x}_0$ from $\boldsymbol{X}$ to be $d_v(\boldsymbol{x}_0)$ $= \|\boldsymbol{v}_0\|_2$. For the purposes of developing a method of assessing the significance of d_v, $D_v(\boldsymbol{x}_0)$ denotes the statistic generating the observed departure $d_v(\boldsymbol{x}_0)$.

2.5. Significance Testing

In this section, a permutation procedure (Good, 1993) is described for testing $H_0 : \boldsymbol{x}_0 \in \mathcal{S}$ versus $H_1 : \boldsymbol{x}_0 \notin \mathcal{S}$ assuming $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are mutually independent. Under H_0, $\boldsymbol{X} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a random sample from $\mathcal{S}$, and the distribution of D_v is estimated by enumerating all possible distinct values of $d_v(\boldsymbol{x}_i), k = 0, 1, \ldots, n$. This is accomplished by constructing all $n + 1$ partitions of $\boldsymbol{X}$ consisting of a singleton set $\{\boldsymbol{x}_k\}$ and $\boldsymbol{X}_k$, and computing the departure of $\boldsymbol{x}_k$ from $\boldsymbol{X}_k$. Let $d_v(\boldsymbol{x}_k)$ denote the kth departure value; then, the empirical distribution function $\widehat{F}_0 = \{d_v(\boldsymbol{x}_0), d_v(\boldsymbol{x}_1), \ldots, d_v(\boldsymbol{x}_{n+1})\}$ estimates the null distribution of $D_v(\boldsymbol{X}_0)$. The number of values in $\widehat{F}_0$ that are greater than or equal to the observed departure $d_v(\boldsymbol{x}_0)$ is the test statistic, and the p-value, or observed significance, level is the proportion of values in $\widehat{F}_0$ greater than or $d_v(\boldsymbol{x}_0)$, i.e.,

$$\widehat{P}[D_v(\boldsymbol{x}_0) \geq d_v(\boldsymbol{x}_0) \mid \boldsymbol{x}_0 \in \mathcal{S}] = \frac{\#\{d_v(\boldsymbol{x}_k) \in \widehat{F}_0 \mid d_v(\boldsymbol{x}_k) \geq d_v(\boldsymbol{x}_0)\}}{n+1}.$$

In this study, this test is used only to provide a quantitative measure of evidence of departure for

comparative purposes, and not for formal testing. Two reasons argue against assigning statistical significance to $\|v_0\|_2$. First, departure analysis is typically carried out for simultaneously for several hundred strata comprising a region of interest, and some provision for controlling experiment-wise Type I·error rates should be available. These methods usually depend on independence among tests, yet departure may be spatially correlated because strata that are close in space tend to be similar with respect to the distribution of PVT's, and because the simulated fires often spread to adjacent strata. In addition, we observed evidence of autocorrelation when observation were generated using a 20 year sampling interval. Thus, the computed values $\widehat{P}[D_v(\boldsymbol{x}_0) \geq d_v(\boldsymbol{x}_0) \mid \boldsymbol{x}_0 \in \mathcal{S}]$ were not used for formal hypothesis testing. For simplicity, we refer to $\widehat{P}[D_v(\boldsymbol{x}_0) \geq d_v(\boldsymbol{x}_0) \mid \boldsymbol{x}_0 \in \mathcal{S}]$ as the *estimated* p-value with the understanding that the estimator is probably biased because the observations $d_v(\boldsymbol{x}_1)$, $\ldots$, $d_v(\boldsymbol{x}_n)$ are not independent.

2.6. Mapping Departure

The primary application of this methodology is via GIS, and so mapping departure is a key step in the analysis of departure. Two approaches are useful; the first is traditional gray-scale mapping of departure, and the second integrates the strength-of-evidence measure obtained from the estimated p-value by mapping three integrated classes, low, medium and high departure. Table 1 shows how the classes are determined from departure and estimated p-value.

Table 1. Departure classes. The departure of the current year from historic conditions is denoted by $d_v(\boldsymbol{x}_0)$ and the estimated p-value is denoted by p.

	Estimated p-value		
Departure	$p<0.01$	$0.01 \leq p < 0.1$	$0.1 \leq p$
$d_v(\boldsymbol{x}_0) < 1/3$	medium	low	low
$1/3 \leq d_v(\boldsymbol{x}_0) < 2/3$	medium	medium	low
$2/3 \leq d_v(\boldsymbol{x}_0)$	high	high	medium

3. An Example

Lands adminstered by the USDA Forest Service in central Utah (Zone 16 in the Forest Service Nationwide Map) are one of two pilot areas used to develop and test the methodology. This region consists of 465 strata, most of which consist of 90030×30 m pixels. The predominant vegetation is montane conifer forests, though there are extensive grass and shrublands, and subalpine forests and alpine meadows. Fire suppression was vigorously pursued in this region during the last century, as is true of nearly all federally administered lands. Figure 2 is a level plot constructed from a typical stratum showing the numbers of observations in each PVT/succession class combination on the horizontal axis and observation year on the vertical axis. Most of the 900 pixels tend to belong to only a few dominant PVT/succession class combinations. In central Utah, these are typically mid-seral and multistrata forest communities. Figure 2 also shows that there are many PVT/successional class combinations that occur infrequently in time and space. Those classes that tend to dominate with respect to pixels counts tend to maintain large counts for extended intervals interrupted by short intervals with small

counts. Long intervals correspond to the development of early and mid-seral forest communities, or established multistrata forest communities, whereas short intervals tend to correspond to early-successional, post-fire nonforest communities. This pattern generally reflects the frequent, low intensity fire disturbance processes of montane conifer forests of the Western United States.

Figure 3 is a gray-scale map of departure for Zone 16, and Figure 4 is a map of departure class, respectively for the central Utah study area. The region consists of approximately 62,308 km^2, maintained in a GIS comprised of 76,924 strata. Table 2 describes the departure distribution over the region by classifying departure and estimated p-value. While Table 2 suggests that the relationship between departure and p-value is not strong because most of the estimated p-values fell in the interval $[0.01,0.1)$, the Pearson correlation coefficient was -0.52.

Table 2. Cross-tabulation of departureand estimated p-value for Zone 16 based on $N=76,924$ observations. The departure of the current year from historic conditions is denoted by $d_v(\boldsymbol{x}_0)$ and the estimated p-value is denoted by p.

	Estimated p-value		
Departure	$p<0.01$	$0.01\leq p<0.1$	$0.1\leq p$
$d_v(\boldsymbol{x}_0)<1/3$	0.00	42.65	0.00
$1/3\leq d_v(\boldsymbol{x}_0)<2/3$	15.92	14.79	10.12
$2/3\leq d_v(\boldsymbol{x}_0)$	0.00	29.15	0.43

Finally, a brief summary of the dimensionality of these data is obtained from a random sample of 256 strata. The first, second and third quantiles for the mean number of columns of the data matrix $\boldsymbol{X}$ was83, 110, and 165, the corresponding quantiles for the rank of $\boldsymbol{X}$ were 80,106 and 158, and the median difference between number of columns and the rank was 5.

4. A Simulation Study

A simulation study was carried out to compare the sensitivity of the departure measures discussed above. This was carried out by combining 200 observations generated from the historic population $\mathcal{P}$ (sampled at 20-year intervals) and atypical observations from a different population, denoted by $\mathcal{P}_0$. Then, the departure of each observation in the combined sample was computed and classified as unusual, or not, depending on the departure value. An observation was classified as unusual if its departure value was larger than the 95th quantile of the sample of departure values. The sensitivity and specificity of a departure measure was quantified by the proportion of observations from $\mathcal{P}_0$ and $\mathcal{P}$, respectively, that were classified as unusual.

The simulation study is similar to those used in studies of outlier detection methods to the extent that a set of unusual observations contaminates a sample, and the objective is to determine the accuracy with which the unusual observations are detected as unusual (and identified as outliers). An important property of outlier detection methods is robustness against swamping. While resistance to swamping is not directly relevant to the comparison of current and historical conditions given that the objective is to determine if a single observation on current

conditions is different from historic conditions, an analysis of resistance to swamping does provide some additional information on the row projection method.

For a particular stratum, atypical observations (from $\mathcal{P}_0$) were obtained by randomly sampling with replacement m observations from the historic sample of n observations, where m was a predetermined fraction of n. Contaminant observations were introduced according to six rates: 1, 5, 10, 15, 20, and 25% of the historical data sample of $n=200$ observations. These additional observations were made atypical by randomly permuting the succession class counts within each PVT class. Specifically, suppose that the p-vector $\boldsymbol{x}_i \in \mathcal{P}$ consists of counts from s PV types with $p_1, \ldots, p_s$ succession classes in each PV type, then the p_j counts from the jth PV type were randomly permuted for each $j = 1, \ldots, s$. Because counts were permuted only within PVT class, observations from $\mathcal{P}$ and $\mathcal{P}_0$ were subject to the same linear constraints, and the rank of the combined data matrix was the same as the historical data set. At the same time, these additional contaminant observations tended to be inconsistent with the process that generated $\mathcal{P}$.

A second simulation scheme was used in which outliers were obtained random sampling m observations from different strata than the stratum being analyzed. Consequently, there were sometimes different PV types were present among the m contaminant observations and the n historic observations, and the rank of the combined data matrix tended to be greater than the historical data matrix. Results from the two schemes were not substantially different, so we report the results only for the first simulation scheme.

The outlier detection methods under study differed with respect to the way in which observation-to-set distances are measured. The distance measures are the row projection error distance d_v, Mahalanobis distance d^r_M for less-than-full-rank covariance matrices, Gnanadesikhan and Kettenring (1972) distance weighting d^r_{GK}, the two distance measures based on the $s = 5$ least important eigenvectors d^s_1 and d^s_2, and the chi-square distance measure d_χ. An observation $\boldsymbol{x}_k$ was identified as an outlier if the distance to the historic data set $\boldsymbol{X}_k$ (after removing $\boldsymbol{x}_k$) was greater than the qth quantile of all such distances $k = 1, \ldots, n$. Two choices of q were used, .90 and .95. Because there was very little difference in the relative performance of the methods between the two quantiles, we present results only for the 95th quantile. Two statistics were computed to assess the sensitivity and specificity of the methods: the detection rate (percent of contaminants identified as outliers), and the mis-identification rate (percent of non-contaminants identified as outliers). By using the 95th quantile as a criterion for outlier detection, 10 observations will be identified as outliers by each method. If the contamination rate is 1%, then 2 observations are contaminants; if these are correctly identified so that the detection rate is 1.00, then 8 additional non-contaminants will be identified as outliers and the misidentification rate will be $8/200=0.04$.

Figures 5 and 6 summarize the results of the simulation analysis. Sensitivity of the row projection method is substantially greater than other methods, particularly when the fraction of contaminants is small, ranging from nearly 1.00 (1% contamination) to 0.22 (25% contamination). While sensitivity declined with increasing contamination rate, specificity of the row projection method was consistently small.

5. Discussion

Simulated data are limited by their dependence on model assumptions. Despite limitations, these data are valuable for assessing *relative* departure across contiguous regions for the purpose of identifying areas with plant communities that are substantially different than historic conditions. Thus, the appropriate use for the methodology discussed herein is for prioritizing landscape units for field visitation and possible treatment. Our recommendation to land managers is to use a two-stage procedure. In the first stage, a subset of all strata within a region with large departure values and small estimated p-values should be identified. The second stage is ground visitation and inspection, and a final determination of management activities.

The methodology presented in this article incorporates several innovations. First, the simulations are spatially explicit; that is, fire disturbance events are simulated by allowing fires to spread as a function of topography, and prevailing wind direction. In contrast, other simulation programs model fire events on each map unit independently, without consideration of fire spread.

Through simulation, historical conditions are quantified by a chronological sequence of successional communities under natural fire regimes. These simulated data are used to compare the historical distribution of plant communities for a landscape unit or a larger landscape. The data used for the comparison are frequencies of occurrence of PVT/successional class combinations, and departure of current from historical conditions is measured by the error incurred when using the historical data to produce the best linear approximation of the current data vector. The simulation study indicates this measure of departure to be substantially more sensitive than other techniques for this purpose. We attribute this sensitivity to the projection operator $\boldsymbol{P}_0$ used to construct the linear approximation. This operator provides a very close approximation of the left-out historic observations for two reasons. First, the historic data are generated by a single stochastic process; consequently, observations are inherently similar. Secondly, $\boldsymbol{P}_0$ provides a very accurate linear approximation because no other linear approximation of the same rank yields a smaller mean square error. In contrast, the outlier detection methods discussed above use information expressed by a subset of all axes of variation extracted from the sample variance matrix $\boldsymbol{D}$. The use of $\boldsymbol{D}$ may be a poor strategy because it is not uncommon that the counts for some PVT/successional classes are predominantly large but with small sample variances; while many other PVT/successional classes are consistently small, also with small variances. Yet, the large count classes are important for determining departure because if none of these frequently occurring classes are well-represented in the current observation, then, current conditions are dissimilar from historic conditions. However, small variances associated with these classes will affect the importance of these classes in determining the departure measure.

6. Acknowledgements

This research was supported in part by funds provided by the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture

7. References

Arno, S. F., D.G. Simmerman, and R. E. Keane. (1985) Forest succession on four habitat types in western Montana. USDA Forest Service General Technical Report INT 177, Ogden, UT., USA.

Arno, S. F., D. J. Parsons, and R. E. Keane. (2000) Mixed-severity fire regimes in the northern Rocky Mountains: consequences of fire exclusion and options for the future. *Wilderness Science in a Time of Change Conference, Volume 5 : Wilderness Ecosystems, Threat, and Management,* USDA Forest Service, Rocky Mountain Research Station, Fort Collins CO..

Botkin, D. B., and H. J. Schenk. (1996) Review and analysis of JABOWA and related forest models and their use in climate change studies. NCASI Technical Bulletin Number 717.

Daubenmire, R. (1966) Vegetation: identification of typal communities. *Science* **151**, 291-298.

DeBano, L. F., D. G. Neary, and P. F. Ffolliott. (1998) *Fire's Effect on Ecosystems.* Wiley and Sons, New York.

Ferry, G. W., R. G. Clark, R. E. Montgomery, R. W. Mutch, W. P. Leenhouts, and G. T. Zimmerman. (1995) Altered fire regimes within fire-adapted ecosystems. USDI, National Biological Service, Washington, DC.

Frost, C. C. (1998) Presettlement fire frequency regimes of the United States: a first approximation. *Tall Timbers Fire Ecology Conference* **20**, 70-81.

Gnanadesikhan, R. Kettenring, J.R. (1972) Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, **28**, 81-124.

GAO. (2002) Severe wildland fires: leadership and accountability needed to reduce risks to communities and resources. *Report to Congressional Requesters GAO-02-259*, US General Accounting Office, Washington DC.

Good, P. (1993) *Permutation Tests.* Springer-Verlag, New York.

Hann, W.J., Jones, J.L., Keane, R.E., Hessburg, P.F., Gravenmier, R.A. (1998) Landscape dynamics. *Journal of Forestry*, **96**, 10-15.

Heinselman, M. L. (1981) Fire and succession in the conifer forests of North America. *In* D. C. West, H. H. Shugart, and D. B. Botkin (eds), *Forest Succession: Concepts and Applications*, Springer-Verlag, New York. pp. 374-406.

Hessburg, P.F., Smith, B.G., Salter, R.B. (1999) Detecting change in forest spatial patterns from reference conditions. *Ecological Appications*, **9**, 1232-1252.

Jackson, D.A., Chen. Y. (2004) Robust principal component analysis and outlier detection with ecological data. *Environmetrics*, **15**, 129-139.

Juan, J., Prieto, F.J. (2001) Using angles to identify concentrated multivariate outliers. *Technometrics*, **43**, 311-322.

Keane, R.E., Long, D.G.,Menakis, J.P., Hann, W.J., Bevins, C.D. (1996) Simulating coarse-scale vegetation dynamics using the Columbia River Basin succession model: CRBSUM. USDA Forest Service Intermountain Research Station, Ogden, UT.

Keane, R.E., Austin, M. Dalman, R., Field, C., Huth, A. Lexer, M., Peters, D., Solomon, A., and Wycoff, P. (2001) Tree mortality in gap models: application to climate change. *Climatic Change*, **51**,509-540.

Keane, R. E., Parsons, R. and Hessburg, P. (2002a) Estimating historical range and variation of landscape patch dynamics: Limitations of the simulation approach. *Ecological Modelling*, **151**, 29-49.

Keane, R. E., Veblen, T. Ryan, K.C., Logan, J., Allen, C. Hawkes, B. (2002) The cascading effects of fire exclusion in the Rocky Mountains. *In* J. B. (ed), *Rocky Mountain Futures: An Ecological Perspective*, Island Press, Washington DC, USA.pp. 133-153.

Kolb, P. F., Adams, D.L., McDonald, G.I.(1998) Impacts of fire exclusion on forest dynamics and processes in central Idaho. *Tall Timbers Fire Ecology Conference*, **20**,911-923.

Landres, P. B., Morgan, P. Swanson, F. J. (1999) Overview and use of natural variability concepts in managing ecological systems. *Ecological Applications*, **9**,1179-1188.

Laverty, L., Williams, J. (2000) Protecting people and sustaining resources in fire-adapted ecosystems - A cohesive strategy. *Forest Service response to GAO Report GAO/RCED, 99-65*, USDA Forest Service, Washington DC.

Mladendoff, D.J., Baker, W.L. (1999) *Spatial modeling of forest landscape change: Approaches and applications*. Cambridge Univ. Press, Cambridge UK, 329 p.

Pena, D., Preito, F.J. (2001) Multivariate outlier detection and robust covariance matrix estimation (with discussion). *Technometrics*, **43**, 286-310.

Penny, K.I., Jolliffe, I.T. (2001) A comparison of multivariate outlier detection methods for clinical laboratory safety data. *The Statistician*, **50**, 295-308.

Pfister, R. D., Arno, S. F. (1980) Classifying forest habitat types based on potential climax vegetation. *Forest Science*, **26**, 52-70.

Rao, C.R. (2004) Statistics: reflections on the past and visions of the future. AMSTAT News, **327**.

Vogel, C.R. (2002) Computational Methods for Inverse Problems. SIAM, Philadelpia.

Waring, R. H., Running, S. W. (1998). *Forest Ecosystems: Analysis at Multiple Scales*, 2nd ed. Academic Press, San Diego.

Wright, H. E., Heinselman, M. L. (1973) Introduction: ecological role of fire. *Quaternary Research*, **3**, 319-328.

8. Figures

Figure 1. Successional pathways for the high elevation subapline fir PVT as dictated by a successional model (Keane et al. 2002). Cover types are shrub/herb (SH), whitebark pine (WP), lodgepole pine (LP), subapline fir (SF). Structural stages are stand initiation (SIN), shrub/grass/forb (SGF), stem exclusion closed (SEC), understory reinitiation (URI), old forest multistrata (OFM), old forest single strata (OFS), and young forest multistrata (YFM).

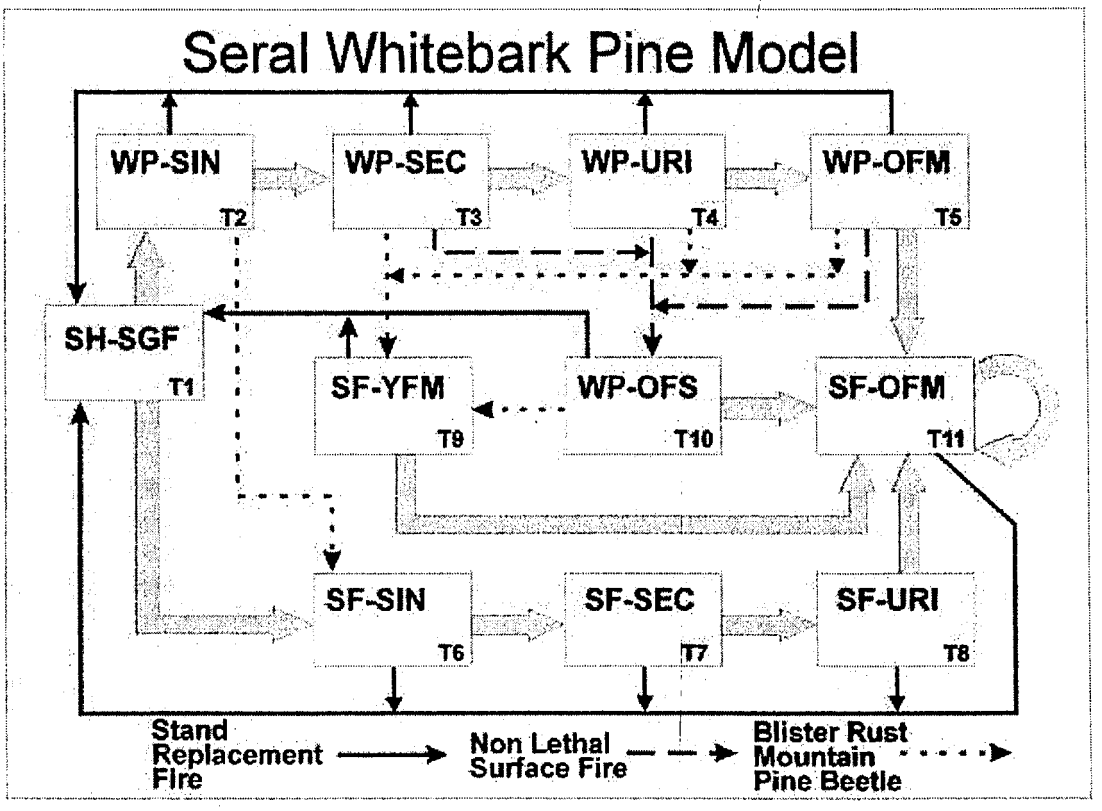

Figure 2. A level plot constructed from a typical stratum showing the numbers of observations in each PVT/succession class combination.

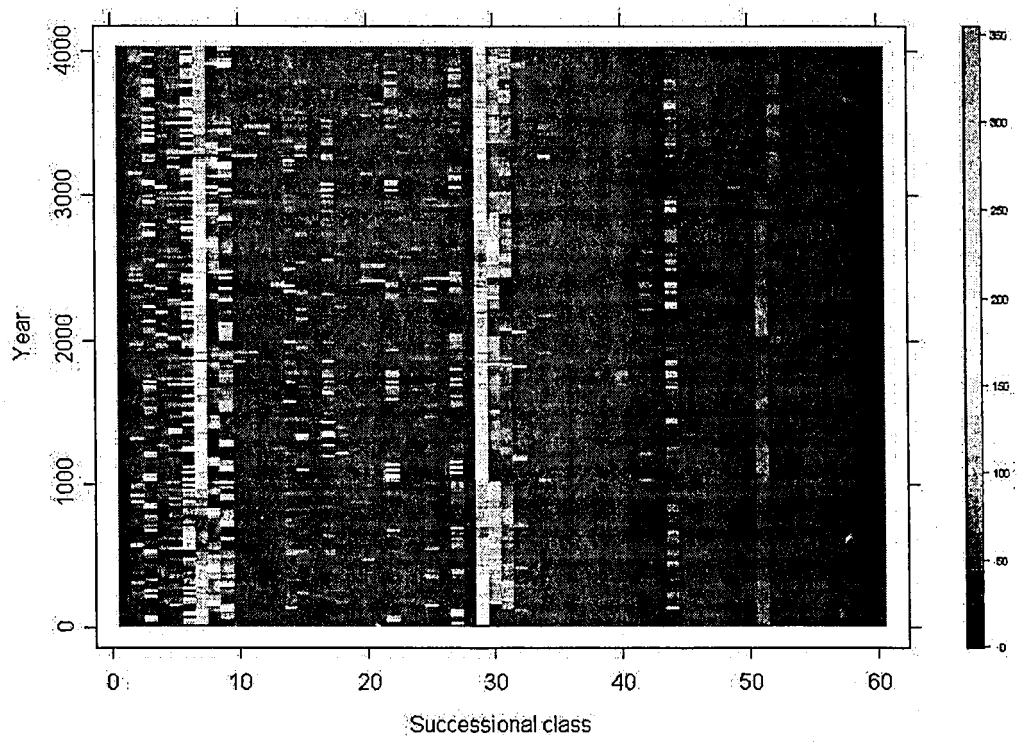

Figure 3. Outlier detection rate plotted against contamination fraction for five measures of departure. Plotted values are averages over 266 strata.

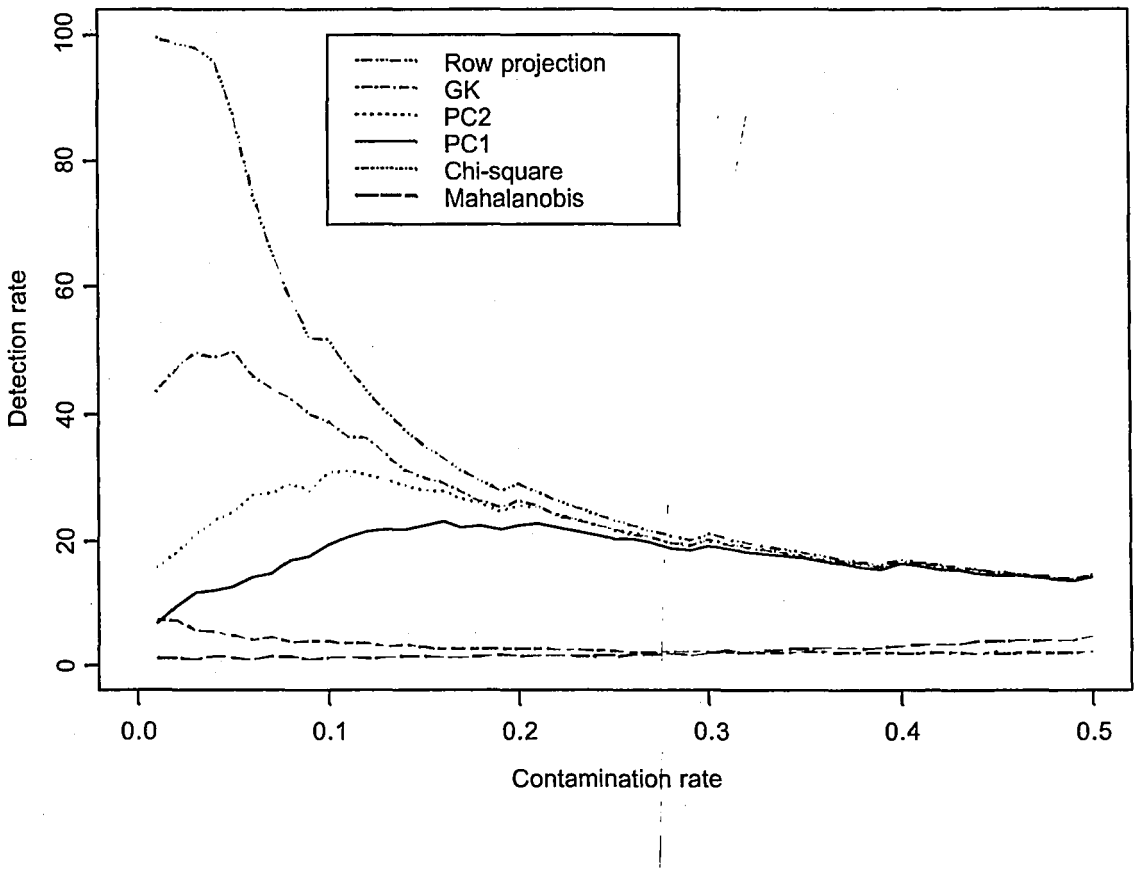

Figure 4. Outlier misidentification rate plotted against contamination fraction for five measures of departure. Plotted values are averages over 256 data sets (strata).

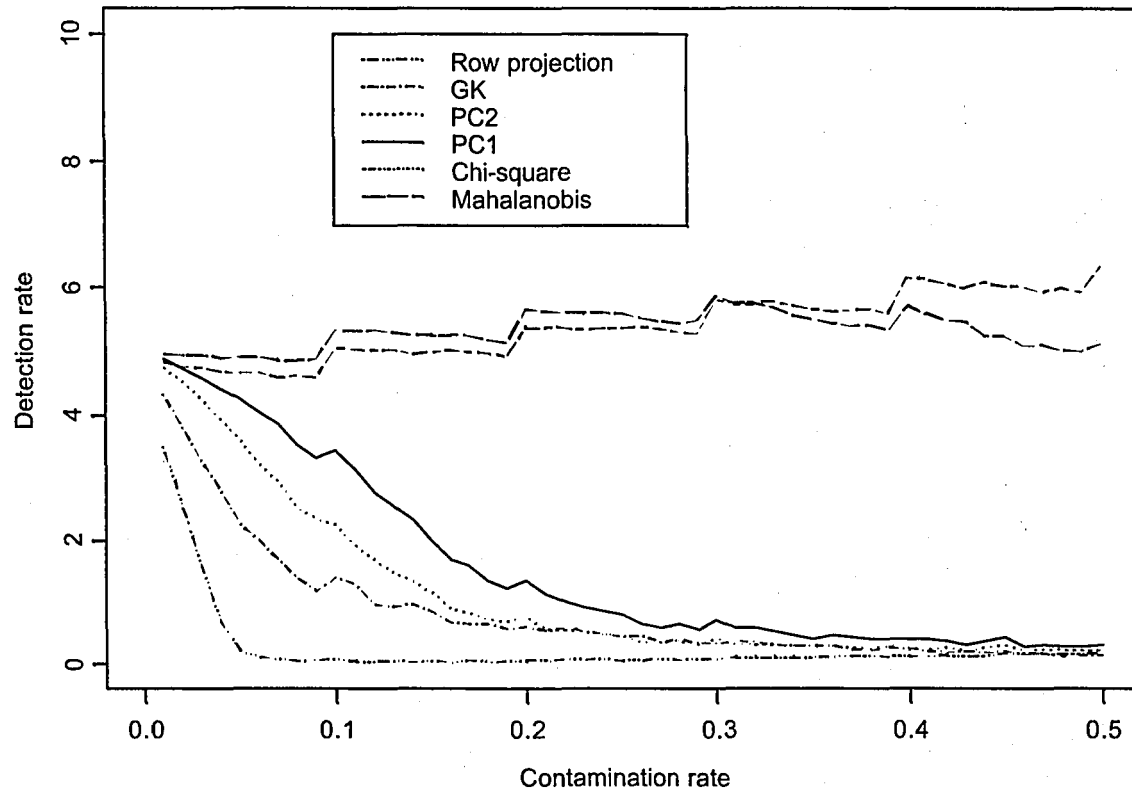

Figure 5. Gray-scale map of departure for central Utah constructed from 76,924 strata.

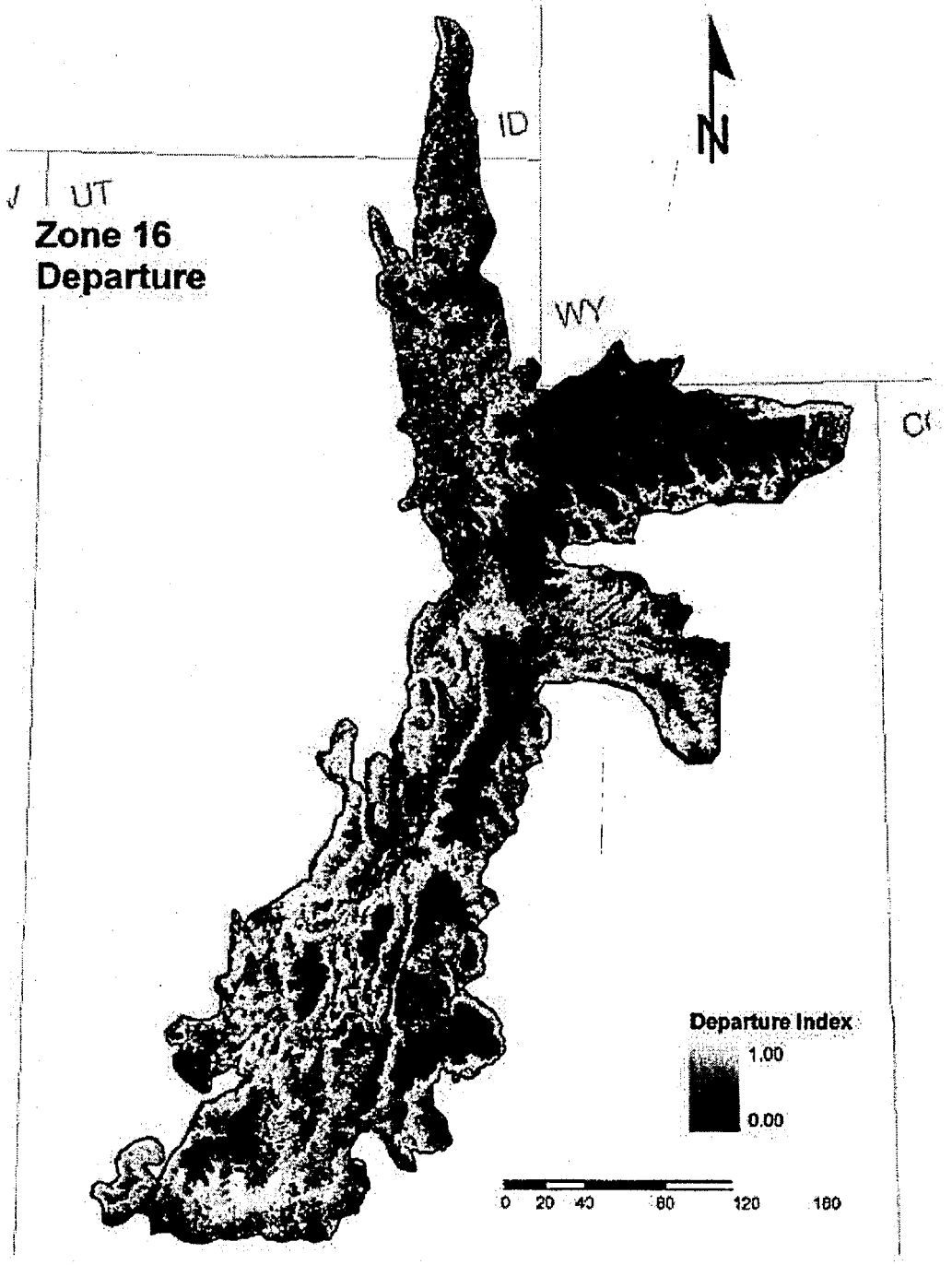

Figure 6. Departure class map of departure for central Utah constructed from 76,924 strata. Class codes are based on both departure and the estimated p-value. Table 2 gives the class code definitions.